
Huang D, Liu Z, Fu X, Blythe PT. [Multimodal Transit Network Design in a Hub-and-Spoke Network Framework](#). *Transportmetrica A: Transport Science* (2018)

DOI link

<https://doi.org/10.1080/23249935.2018.1428234>

ePrints link

<http://eprint.ncl.ac.uk/245493>

Date deposited

04/04/2018

Embargo release date

16/01/2019

Copyright

This is an Accepted Manuscript of an article published by Taylor & Francis in *Transportmetrica A: Transport Science* on 16/01/2018, available online:
<https://doi.org/10.1080/23249935.2018.1428234>

Multimodal Transit Network Design in a Hub-and-Spoke Network Framework

Di Huang^a, Zhiyuan Liu^{a,*}, Xiao Fu^b, Philip T Blythe^c

^a*Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing, Jiangsu 210096, China.*

^b*School of Transportation, Southeast University, Nanjing, Jiangsu 210096, China.*

^c*Transport Operations Research Group, School of Civil Engineering and Geosciences, Newcastle University, UK*

Email addresses: zhiyuanl@seu.edu.cn (Z. Liu), dhuang2@seu.edu.cn (D. Huang), fuxiao@seu.edu.cn (X. Fu), phil.blythe@ncl.ac.uk (P. T. Blythe)

Published in: Transportmetrica A: Transport Science, 2018

DOI: 10.1080/23249935.2018.1428234

To link to this article: <https://doi.org/10.1080/23249935.2018.1428234>

Abstract

This study focuses on the reconfiguration of bus services in an urban area with a newly constructed rail system. A hub-and-spoke network framework is introduced for the multimodal transit system. The rail services are taken as the backbone, and the bus services (main bus lines and feeder bus lines) are reconfigured to better integrate with rail services forming the core of a Mobility as a Service (MaaS). A cluster-based approach is used for selecting hubs from rail stations. Bus stops are taken as non-hub nodes. Main bus lines are designed based on a heuristic line generation approach and feeder bus lines are developed by solving a traveling salesman problem. A bi-level programming model is proposed to determine frequencies of each mode and addressed by the artificial bee colony algorithm. The effectiveness of the proposed methods is illustrated by numerical examples and applied to the Mandl's benchmark compared with existing studies.

Keywords: multimodal transit network design; hub-and-spoke network; cluster-based

hub location problem; bi-level programming problem.

1. Introduction

A multimodal transit system has become an effective method to meet the various types of travel demand in metropolitan areas. As the backbone of the multimodal transit network, the rail system is fast expanded in many cities due to its rapid service, regular schedule and adequate capacity providing inherent convenience to various types of trips (long-, middle- and short-distance trips). It is widely recognized that the newly constructed rail services have significant impacts on the existing passenger demand distribution of bus lines (Li et al. 2009; Li et al. 2011; Schöbel 2012; Saidi et al. 2017). Thus, the reconfiguration of the existing bus network (both, in terms of, the spatial line layout and the frequency design) is necessary to integrate efficiently with the new rail lines to improve the overall accessibility of the resultant multimodal public transit system.

A hierarchical urban multimodal transit system could be constructed on three levels (see Fig. 1(a)): i. The rail system acts as the backbone of the multimodal transit network and provides rapid and long-distance services; ii. The main bus lines connect rail stations that have high passenger volume but are not directly connected by rail lines to provide mid-distance services; and iii. The ad hoc feeder lines serve as the supplementation for rail and main bus lines by delivering first/last mile services between rail stations and passengers' origins/destinations.

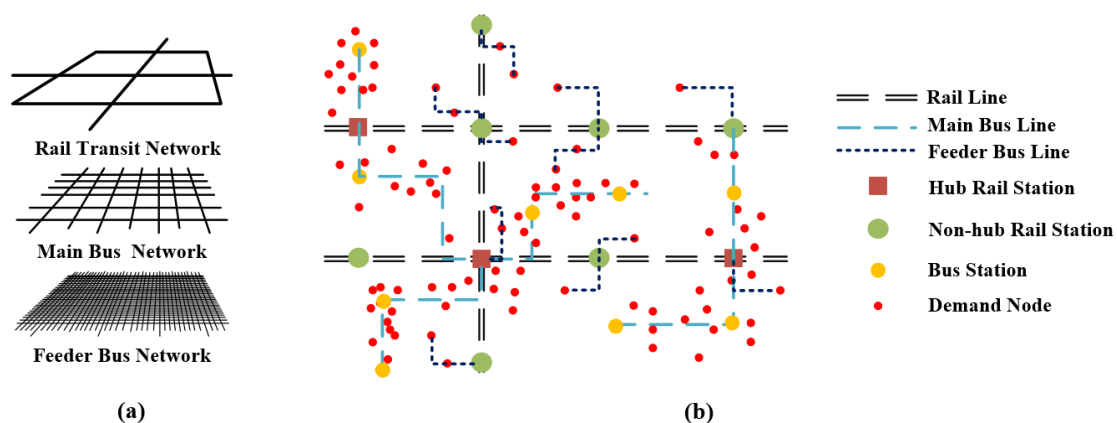


Fig. 1 Illustration of the multimodal transit network.

In the multimodal urban transit system, demands of multiple transit travel mode are interrelated, where a coordinated network framework may offer reasonable combinations of various design decisions (Farahani et al. 2013) rather than a fully integrated and seamless MaaS operation (Blythe 2016). To enhance the connections and interactions within a multi-level transit network, the concept of hub-and-spoke network framework is adopted to deal with the reconfiguration of the existing bus network (main bus lines and feeder bus lines) in the context that a new rail/subway line is newly constructed. Passenger flows are mainly consolidated and encouraged to transfer to rail lines in hubs, which offers a solution to the hierarchical transit network by accommodating multilevel trip demands.

The hub-and-spoke framework was initially developed in the telecommunication industry. But now it has been widely applied to many areas including maritime shipping, air transportation, freight logistics and public transport. This network framework outperforms the traditional separately designed transit network as it concentrates the passenger flow on the rail lines, resulting in an efficient transit capacity sharing and fleet management. Network operators can benefit from the economies of scale by increasing the utilization of network resources (Gelareh 2008; Gelareh and Nickel 2011; Farahani et al. 2013a).

In the hub-and-spoke network framework, rail stations with higher trip attraction/production are termed as hubs and bus stations are non-hub nodes (see Fig. 1(b)). Three steps/tasks are fulfilled to solve the multimodal transit network design problem (MTNDP). The first step is the hub identification. Some of the rail stations are selected as hubs according to the nearby passenger demand density. The second step is the transit line design. The candidate main bus lines and feeder bus lines are generated in a coordinated configuration in the network. The third step is the bus line and frequency deployments, in which we can obtain optimal operating parameters.

With the help of Advanced Traveller Information Systems (ATIS) and Intelligent Transport Systems (ITS), transit agencies and operators have increasingly adopted automatic passenger counter and automatic vehicle location systems to tackle both the long-term strategic planning and real-time transit operation (Chang et al. 2010; Bordagaray et al. 2014; Nuzzolo and Comi 2016). Additionally, the large amount of

data gives more observations of passenger demand in space and time than any other means of data collection during the conduction of strategic planning and real-time operation of the transit network.

1.1 Literature review

The MTNDP intends to create a reliable public transport system by finding a set of hierarchical transit lines and setting corresponding tactical decisions (e.g., line service frequency, fleet size, timetable, etc.) (Bagloee and Ceder 2011; Cipriani et al. 2012; Farahani et al. 2013b; Arbex and da Cunha 2015, among many others). At the strategic level, long-term decisions are made regarding the design of road and multimodal transit networks, whereas the optimal operating parameters are designed at the tactical level. Furthermore, the travel behavior of users concerning transit lines and frequencies is modeled by a transit assignment sub-model, which is necessary to measure the performance of the network design strategy (Cancela et al. 2015). Hence, the MTNDP is usually modeled as a bi-level programming problem (BLPP) (Szeto and Jiang 2014; Szeto et al. 2015; Szeto and Wang 2016), where the upper-level problem optimizes the transit network based on feasibility constraints by the planner, and the lower-level problem describes the response of passengers to the designed scenarios by the upper-level.

A typical hub-and-spoke network firstly proposed by O'Kelly (1986) is depicted in Fig. 2(a). Rather than directly connecting any pair of nodes, each non-hub node must be allocated to just one hub node and hence is termed as the single allocation *p-hub location problem (p-HLP)*. Then, several restrictions are relaxed in some studies with proper fitness to different conditions. Campbell (1996) extended the *p-HLP* by introducing the multiple allocation policy. As every non-hub nodes could be allocated to more than one hub node, this model is named as *multiple allocation p-HLP*. Aykin (1995) further relaxed the restriction on hub service by allowing non-stop service (as shown in Fig. 2(b)), which means that spoke nodes can be connected directly without using any hub.

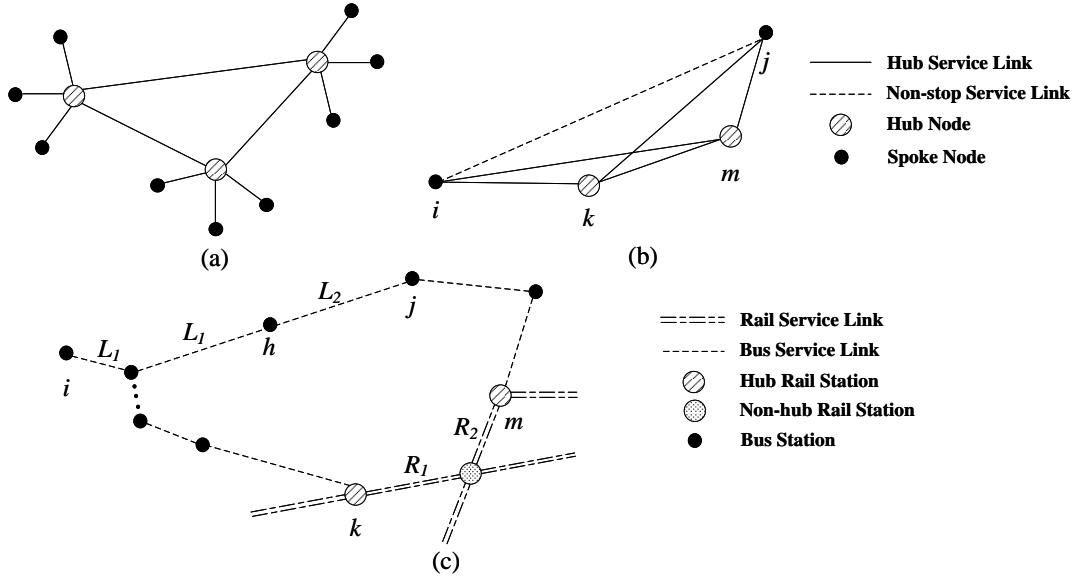


Fig. 2 Illustration of hub-and-spoke network framework: (a) p -HLP (O'Kelly 1986), (b) *multiple allocation* p -HLP under the non-restrictive policy (Aykin 1995), and (c) extended transit network.

Comparing with the hub-and-spoke frameworks applied in the areas of maritime and air transportation (Sang et al., 2014; Yang, 2010), the design of hub-and-spoke multimodal transit network is becoming increasingly complex resulting from the competition and cooperation between different transit modes. Nickel et al. (2001) proposed the first hub-and-spoke transit network. In their study, the rail lines are represented as hub links while the bus lines as non-hubs. Bagloee and Ceder (2011) adopted a multi-phase planning procedure from the perspective of transit authorities. de Sá et al. (2015) proposed the hub line location problem of hub networks with multiple lines, in which the location of a set of hub facilities connected by transit lines is considered.

The concept of the hub-and-spoke network framework has been widely adopted in the MTNDP because of its service efficiency and the economy of scale, which can be divided into discrete (Nickel et al. 2001; Gelareh and Nickel 2008; Wang and Chen, 2012; Yu et al. 2013) and continuous design models (Daganzo 2010; Badia et al. 2014). In the discrete HLP, the hubs are located in a given subset of data points in the service area, which can be efficiently modeled in mathematical formulations such as the mixed integer programming (Farahani et al. 2013a). Nickel et al. (2001) proposed the first mathematical model for the application of the HLP in public transit network. Several

assumptions for the HLP are relaxed for the adaption to public transport planning, such as the allowance of incompletely inter-hub connections. Gelareh and Nickel (2008) proposed a novel mixed integer programming considering the real passengers' willingness in hubs. Yu et al. (2013) developed a two-phase approach to deal with the HLP in public transit networks. However, to the best of the authors' knowledge, the study of the hub-and-spoke network incorporating a rail network has not received any research attention. In practice, the construction of the rail lines would inevitably result in the perturbation of passenger distribution in the existing transit network. Despite its practical significance, few of the existing studies have considered the reconfiguration of the transit network. In this regard, the methodology of deploying the main bus lines on the basis of the rail network is necessary, while the hub-and-spoke network framework can efficiently solve this problem. As shown in Fig. 2(c), a subset of rail stations is selected as hubs, which can be connected to rail and main bus lines conjointly. Meanwhile, non-hub rail stations are only connected with rail lines. The non-hub nodes exteriorly located around the rail system are taken as main bus lines. The feeder bus services are provided at both hub and non-hub rail stations.

In the planar or continuous version of HLP, the hub can be located anywhere in the service area (Iyigun 2013). O'Kelly (1986) was the first one who studied the planar HLP by the clustering technique. Klincewicz (1991) proposed a two-phase clustering-based heuristic algorithm aiming to improve the efficiency of large networks, where the first phase is clustering and the second is to determine the hub location for each cluster. Sung and Jin (2001) introduced a new cluster-based HLP where demand nodes are partitioned into a predetermined number of clusters and only one node in each cluster is selected as a hub by the k-means algorithm. However, because of the uncertainty of the passenger demand distribution, the determination of the number of clusters before clustering is thus difficult to address, which is still an open question in the literature and tackled in this paper.

Table 1 A summary of MTNDP in the hub-and-spoke network framework.

Reference	Objective	Decision	Solution method	Efficiency (# of nodes)	# of hubs
-----------	-----------	----------	--------------------	----------------------------	-----------

Nickel et al. (2001)	Min. total system cost	Discrete	Heuristic	10	6
Bagloee and Ceder (2011)	Min. total saved generalized time	Discrete	Heuristic	13,487	-
de Sá et al. (2015)	Min. total weighted travel time	Discrete	Benders- branch-and-cut and Heuristic	40	-
Gelareh and Nickel (2008)	Min. total construction cost	Discrete	Exact	30	-
Wang and Chen (2012)	Max. operating efficiency	Discrete	Heuristic	100	12
Yu et al. (2013)	Max. served population	Discrete	Heuristic	3000	4
	Min. overlapped served population among the hubs				
	Min. construction cost				
Daganzo (2010)	-	Continuous	Approximation	-	-
Badia et al. (2014)	Min. total network costs	Continuous	Grid search	-	-

1.2 Objectives and contributions

The main objective of this study is to propose a three-stage transit network design method on the newly constructed rail system in the hub-and-spoke framework. The passenger demand node is assumed to be randomly scattered in the service area. Clustering the passenger nodes into zones is the initial work of the HLP. As aforementioned, in the previous studies of the cluster-based HLP, the number of clusters is predetermined, which is unrealistic in practice. Thus, in the first stage, a novel clustering algorithm is proposed to determine the number and locations of the hubs endogenously.

In the second stage, a multimodal transit network is designed hierarchically based on the rail system, including main and feeder bus services, resulting in a two-phase heuristic candidate line generation approach. In order to improve the efficiency of the transit system, the views of both operator and passenger are taken into consideration.

In the third stage, a BLPP is adopted to describe the decision process of optimal lines and their corresponding service frequencies, which influences both the passenger's travel cost and operator's cost. Thus, the objective function of the upper-level problem is defined as the minimization of total social cost which is the sum of the operator's cost and the user's cost. In the lower-level problem, the optimal strategy-based transit assignment method proposed by Spiess and Florian (1989) is adopted to describe the passenger's route choice behavior. Due to the inherent complexity and NP-hardness of BLPP, a heuristic algorithm is adopted to solve the optimization model.

To sum up, the contributions of this study can be summarized in three aspects. First, the study appears to be the first devoted exclusively to the design of multimodal hub-and-spoke transit network based on a newly constructed rail network. The results would provide guidance on the reconfiguration of bus lines after new rail lines are constructed. The second contribution is the proposal of a three-stage multimodal transit network design method. The HLP and MTNDP are solved in the first two stages and a bi-level optimization model is formulated in the last stage to obtain the optimal tactical decision variables. The third contribution of this study is the faster and more flexible clustering algorithm proposed for HLP, which can intuitively and efficiently generate the number and locations of cluster centers.

The paper is organized as follows: a three-stage cluster-based multimodal transit network design method is discussed in the next three sections. Section 5 describes our proposed Artificial Bee Colony (ABC) algorithm. A numerical example is presented in Section 6. Finally, conclusions are drawn, together with suggestions for further studies.

2. Stage 1: Clustering of the hub location

For a newly constructed rail network, only a few stations can be taken as hubs because of the budget constraint. Considering the scattered distribution of passenger demand in a planar region, a clustering procedure is needed to select the rail stations located in the areas with higher passenger demand. The k-means algorithm has been widely adopted in the planar HLPs (O'Kelly 1992; Sung et al. 2000). As one of the essential inputs, the number of clusters should be predetermined which has significant impacts on the clustering result. In this regard, a density-based clustering algorithm is firstly applied in the planar HLP, where the number of clusters is not required which

could be detected automatically according to the density of data points (Kriegel 2011; Rodriguez and Laio 2014).

In the literature, the node fulfilling the following two criteria will be selected as hubs: (i) having a relatively higher passenger demand, and (ii) distant from any other node with higher demand density. Accordingly, for each demand node i , two quantities are needed: the local density ρ_i , and the distance φ_i between node i and other points with higher density. Both quantities are calculated considering the Euclidean distance d_{ij} between demand nodes i and j , which should satisfy the triangular inequality as well. According to Rodriguez and Laio (2014), the local density ρ_i can be defined as

$$\rho_i = \sum_j \phi(d_{ij} - d_0) \quad (1)$$

where $\phi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases}$, and $x = d_{ij} - d_0$. d_0 denotes the radius of the catchment area of a transit terminal. The local density ρ_i can be interpreted as the number of demand nodes that are closer than d_0 to node i . The relative distance of node i (denoted as φ_i) is defined as the minimum distance from node i to any other node with the higher local density:

$$\varphi_i = \min_{j: \rho_j > \rho_i} (d_{ij}) \quad (2)$$

Note that two extreme cases should be considered additionally: (i) if node k has the highest local density, it is natural to take $\varphi_k = \max_j (d_{kj})$; (ii) if the local densities of node i and j are equal (i.e. $\rho_i = \rho_j$) and the Euclidean distance between the two nodes are relatively small (i.e. $d_{ij} \leq d'_0$), then we only retain the node with larger relative distance and set the smaller φ as zero.

The proposed clustering algorithm is illustrated by a small example in Fig. 3. Fig. 3(a) shows the distribution of 60 random points in a two-dimensional space, and Fig. 3(b) shows the scatter plot of ρ_i in terms of φ_i for each point, which is named as the decision graph of clustering (Rodriguez and Laio 2014). Fig. 3(b) illustrates that

points A and B have relatively higher ρ and φ , which are termed as cluster centers and selected in Fig. 3(a). Then, all points in the catchment area of a cluster center are assigned to this center, and these points are termed as a cluster, or named as a Traffic Analysis Zone (TAZ). Note that the points with relatively high φ and low ρ (located at the lower left corner of Fig. 3(b)) are termed as outliers, which will not be assigned to any cluster.

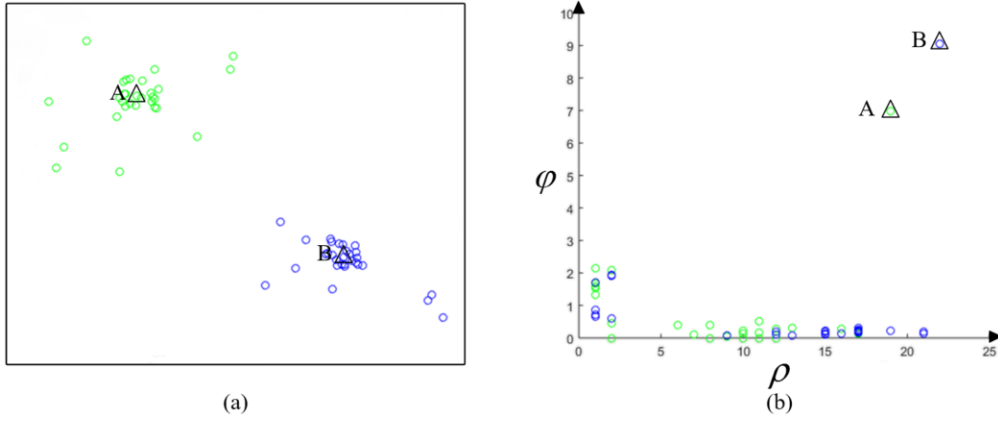


Fig. 3 Illustration of the clustering algorithm: (a) node distribution; (b) decision graph for the data in (a).

As one of the important inputs of clustering, the radius of the catchment area of transit terminals affects the results of clustering, especially the relative location of cluster centers. In real multimodal transit networks, passengers will be assigned to a transit terminal only when they are located within the catchment area of this terminal. Generally, the catchment area of a rail station is considered to be larger than that of a bus station because passengers may take the feeder service as the connector between origins and rail stations. In this regard, the clustering procedure is conducted in two phases for rail and main bus stations respectively.

In Phase 1, the number and the location of hub rail stations are selected. The rail stations located in the catchment area of a cluster center are established as hubs. Although the number of hub rail stations is intuitively determined by the decision graph (refer to Fig. 3(b)), the maximum number of hubs should be limited with taking into account the construction cost of hub facilities. According to the decision graph of the

clustering procedure, only the nodes located at the top right corner of the decision graph with both high density and long distance can be selected as hubs. Therefore, for each node, a weighted index (denoted as w_i) is introduced to express the “quality” of the node, which is defined as a weighted sum of ρ_i and φ_i :

$$\eta_i = \mu_1 \cdot \rho_i + \mu_2 \cdot \varphi_i \quad (3)$$

where μ_1 and μ_2 are the weights of ρ_i and φ_i . Without loss of generality, the two predetermined weights in Eq. (3) are set as 1.0 for the subsequent discussions. The index η_i can be adopted as the criterion of hub nodes selection.

In Phase 2, a set of candidate main bus stops is generated by the same clustering algorithm. In this phase, the cluster center can be treated as the “centroid” of a TAZ and all passengers in this TAZ are allocated to the centroid. All centroids are assumed to be fully connected. Additionally, the outlier passengers with higher φ and lower ρ can be considered as TAZs composed of a single point, which would be picked up by feeder bus.

3. Stage 2: Multimodal transit network design

In a multimodal transit network, the rail system is usually considered as the skeleton, based on which other transit services are designed hierarchically (Cipriani et al. 2012). In this stage, the main and feeder bus lines are generated based on the candidate bus stops and rail stations selected in Stage 1. Due to the nonlinearity and non-convexity of the MTNDP, a two-phase heuristic line generation procedure is proposed in this stage, wherein the main and feeder bus lines are designed separately. In the first phase, two sets of feasible main bus lines are built on different criteria from the perspectives of user’s and operator’s respectively. The first set (termed as Set A) is composed of direct lines connecting origin–destination (OD) pairs with higher demands to accommodate the user’s inclination as to which line choice they would prefer. The second set (termed as Set B) is developed from the operator’s point of view with higher operation efficiency. Then the lines in both Set A and B are filtered by several realistic feasibility constraints and stored in the third set (termed as Set C). The details of each

process in Fig. 4 are elaborated in the remainder of this section. In the second phase, the feeder bus line is generated according to the actual demand at each rail station, which can be formulated by the traveler salesman problem (TSP) (Chandra and Quadrifoglio 2013).

Phase 1: Generating main bus lines

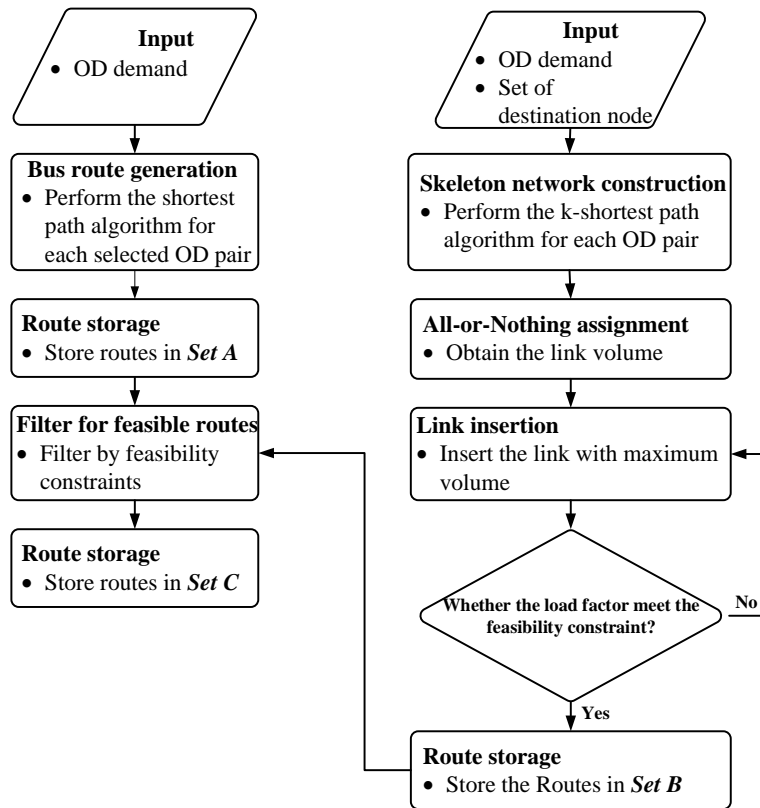


Fig. 4 Flowchart of the proposed main bus generation method.

3.1 Phase 1: Generating main bus lines

3.1.1 Set A lines generation

In public transit systems, the passenger intends to choose the path with fewer transfers and prefers direct trips between the main bus terminals (Ceder 2007; Yan et al. 2013). Hence, the lines in Set A are generated by the shortest path algorithm that connects all OD pairs with higher demands that is larger than a predetermined minimum value, q_{\min} .

3.1.2 Set B lines generation

From the operator's point of view, the construction of direct lines requires a large number of lines (to meet various OD demand) which would result in a high operation cost and a less efficient transit network (Cancela et al. 2015). Thus, the operator's design strategy is to generate the bus lines with higher operation efficiency instead of the direct line which would lead to an increase of transfers compared with those of Set A. Thus, a heuristic line generation method with link insertion is adopted that considers the link passenger flow obtained by a demand assignment subroutine (Cipriani et al. 2012).

Given the OD demand matrix, the destination terminals of the main bus lines can be selected on the basis of passenger demand and other realistic elements. The terminal selection method based on the node gravity proposed by Bagloee and Ceder (2011) is adopted that considers both passenger production/attraction and relative distance. For each terminal, the incoming link is selected if the link volume is greater than a predetermined minimum value, v_{\min} . The line is then expanded by adding other links from the current node of this line. Consequently, this insertion should be terminated if a specific constraint is violated. A tentative frequency, f_l , is adopted to calculate the load factor of each line,

$$f_l = \frac{v_{\max}^l}{LF_{\max}^l \cdot k_{cap}^l} \quad (4)$$

where v_{\max}^l is the largest link volume of line l . LF_{\max}^l is the maximum load factor of line l . k_{cap}^l is the bus capacity. The link insertion procedure will terminate when the load factor of this line is smaller than a predetermined threshold, LF_{\min}^l .

The line generation procedures are summarized as follows:

Step 1. Select main routes destination terminals;

Step 2. All-or-nothing demand assignment between each selected OD pair;

Step 3. Calculate tentative line frequency;

Step 4. Select the entering link of the reached node with the highest volume;

Step 5. If the inserting link satisfies the feasibility constraints, insert the entering link into the route, then go to *Step 4*; otherwise, store the current line in Set B.

3.1.3 Set C lines generation

In the last two sections, two complementary sets of lines are generated from both the operator's and passenger's points of view, which should be checked for practical feasibility constraints and stored in Set C. First, in the hub-and-spoke framework, the passenger flows are encouraged to consolidate in hub rail stations and transported to various destinations through rail lines. Thus, each bus line should pass at least one hub rail station. Second, the length of the main bus line should be limited in a feasible range to meet the demand of mid-distance trips. Third, the distance between each bus stations, called the interstation spacing, should be larger than a predetermined value to reduce the overlapping of catchment areas between adjacent stations (Baaj and Mahmassani 1995). In summary, the Set C is composed of the bus lines that meet the following criteria:

- (i) Hub passage constraint: every bus line passes at least one hub rail station;
- (ii) Total length constraint: the length of a bus line is within a threshold;
- (iii) Minimum interstation spacing constraint: the interstation spacing is larger than a predetermined threshold, which is related to the actual size and specific characteristics of the transit network;
- (iv) The overlapping of bus lines is verified by comparing the length of each line. If the overlap length of two lines exceeds a maximum length value, the shorter one is discarded.

If all the constraints are complied with, the feasible main bus lines are stored in Set C as the input data for the third stage of the MTNDP.

3.2 Phase 2: Generating feeder bus lines

The feeder bus service has been considered as an effective approach to solving the first/last mile problem between residential areas and rail stations (Bagloee and Ceder 2011). The feeder bus line design problem is considered as a TSP in the

catchment area of a station (Chandra and Quadrioglio 2013). It is assumed that the proportion of passenger demand of using the feeder bus service at rail station k is α_k , $0 < \alpha_k < 1$. Thus, this problem is actually a TSP-tour with $\alpha_k \cdot q_k$ nodes, where q_k is the passenger demand of rail station k .

4. Stage 3: Transit network optimization

As mentioned in the Introduction, the decision concerning tactical variables (e.g., line layout, frequency, fleet size, etc.) is another necessary process of the MTNDP. In what follows, the optimal or near optimal sub-set of lines together with their frequencies are obtained through an optimization model.

When designing a transit network, two stakeholders with conflict objectives have to be addressed, including the transit company (operator) and the passenger (user). First, the transit company intends to minimize their operation cost which is a function of the frequency and the roundtrip travel time. Second, the passengers seek to minimize their total travel cost comprising the sum of in-vehicle travel time, waiting time and transfer time. Additionally, the relationship between the transit company and the user is usually described as a leader-follower game (Farahani et al. 2013b; Szeto and Jiang 2014). As the leader, the transit company plans and operates the transit network, while the passengers, as the follower, react based on the decisions of the leader, which can be formulated by a bi-level optimization model.

4.1 Upper-level problem

Denote the multimodal transit network as $G = (N, A)$, where N is the set of nodes and A is the set of directed links. Let D be the set of all bus station nodes (including both main bus and feeder bus stations), and P be the set of all rail stations, thus $D \subseteq N$, $P \subseteq N$, and $N = D \cup P$. The sets of hub and non-hub rail stations are denoted as P_h and P_n . The sets of all lines of rail and main bus are denoted by L_R and L_B respectively. The feeder bus line is identified by the rail station it serves, which can be denoted by L_F^P . Let W be the set of OD pairs, $w \in W$.

In a multimodal transit network, the passenger mode choices can be categorized

into three categories: the combination of rail and bus services (i.e. multimodal), the bus service only, and the rail service only. For the first group, the total travel time is equal to the sum of: (i) the travel time from origin i to the first visited rail station k , (ii) the access time at station k , (iii) the in-vehicle travel time between station k and the last rail station m within the rail transit system, (iv) the transfer time within the rail transit system, (v) the exit time at station m , and (vi) the travel time from station m to destination node j . For the second group, the total travel time is the sum of: (i) the travel time from origin i to the transfer station h , (ii) the transfer time at station h , and (iii) the travel time from station h to destination j . For the third group, the total travel time is the sum of: (i) the time to access a rail line at station k , (ii) the travel time between the first rail station k and the last rail station m , (iii) the exit time to leave the rail line at the last visited rail station m , and (iv) the transfer time for changing lines at one or more rail stations.

Based on the above discussions, several notations are introduced and explained as follows. The in-vehicle travel time of an arc $(i, j) \in A$ is represented by t_{ij} , $t_{ij} > 0$. The access and exit times to enter the first hub rail station k and leave the last hub m are defined by $\varphi_k^a > 0$ and $\varphi_m^e > 0$, respectively. In this study, both φ_k^a and φ_m^e are assumed to be constant parameters. The time of transfer is defined as the combined waiting time at a node n :

$$\varphi_n^t = \frac{1}{\sum_{l \in L_n} f_l} \quad (5)$$

where f_l is the frequency of line l . L_n is the set of all transit lines stop at transfer n .

Hence, the total travel time of each mode between an OD pair $w \in W$ can be calculated as follows:

$$\text{Multimodal: } T_1^w = \sum_i \sum_j t_{ij} + \sum_k \sum_m (\varphi_k^a + \varphi_k^e + \varphi_m^e) + \sum_h \varphi_h^t \quad (6)$$

$$\text{Bus only: } T_2^w = \sum_i \sum_j t_{ij} + \sum_h \varphi_h^t \quad (7)$$

$$\text{Rail only: } T_3^w = \sum_k \sum_m (\theta_k^w + \theta_k^b + \theta_m^b) \quad (8)$$

The passenger's mode choice behavior is estimated by a probabilistic modal split model, which is computed by applying multinomial logit model (Eq. (9)). The utility function is calculated by the total travel time according to Eqs. (6)-(8).

$$P_i^w = \frac{\exp(-\theta T_i^w)}{\sum_i \exp(-\theta T_i^w)}, i = 1, 2, 3 \quad (9)$$

where P_i^w is the passenger rate that chooses multimodal, bus service only and rail service only in OD pair w , respectively.

Hence, the total travel cost of all passengers can be expressed as:

$$z_1(\mathbf{l}, \mathbf{f}, \mathbf{q}) = \sum_w q_w \cdot (P_1^w \cdot T_1^w + P_2^w \cdot T_2^w + P_3^w \cdot T_3^w) \quad (10)$$

where \mathbf{l} is the vector of main and feeder bus lines. \mathbf{f} is the vector of line service frequency. \mathbf{q} is the vector of passenger volume. q_w is the passenger demand between OD pair w . P_1^w , P_2^w and P_3^w are obtained in the lower-level transit assignment problem.

The operator's costs depend on the total bus travel distance and time in terms of bus line length and service frequency, respectively. Thus, the operation cost can be calculated by:

$$z_2(\mathbf{l}, \mathbf{f}) = C_R \cdot \sum_{m \in L_R} LL_R^m \cdot f_R^m + C_B \cdot \sum_{n \in L_B} LL_B^n \cdot f_B^n + C_F \cdot \sum_{k \in L_F} LL_F^k \cdot f_F^k \quad (11)$$

where C_R , C_B and C_F represent the per-kilometer operating cost, and all in \$. LL_R^m , LL_B^n , and LL_F^k denote the line length.

Thus, the objective function of the upper-level problem z is

$$\min_{\mathbf{l}, \mathbf{f}} z(\mathbf{l}, \mathbf{f}, \mathbf{q}) = \lambda \cdot z_1(\mathbf{l}, \mathbf{f}, \mathbf{q}) + z_2(\mathbf{l}, \mathbf{f}) \quad (12)$$

where λ is the value of time in \$/h. The constraints for Eq. (12) are then introduced as follows. First, the fleet size used in each transit mode cannot exceed the available fleet size, thus

$$\sum_{m=1}^{L_R^{\max}} 2RT_R^m \cdot f_R^m \leq W_R, m \in L_R \quad (13)$$

$$\sum_{n=1}^{L_B^{\max}} 2RT_B^n \cdot f_B^n \leq W_B, n \in L_B \quad (14)$$

$$\sum_{k=1}^{L_F^{\max}} 2RT_F^k \cdot f_F^k \leq W_F, k \in L_F \quad (15)$$

where W_R , W_B , and W_F are the maximum fleet size of rail, main bus and feeder bus. The capacity constraint is formulated to ensure enough line capacity to meet the total passenger demand

$$\sum_{l=1}^{L_{\max}} f_l k_{cap}^l \geq \sum_{w \in W} q_w \quad (16)$$

where k_{cap}^l is the capacity of line l .

The feasibility constraint defines both the minimum and maximum values for frequencies

$$f_{\min} \leq f_l \leq f_{\max}, l \in L. \quad (17)$$

4.2 Lower-level problem

In the lower-level problem, a static strategy-based transit assignment model is adopted to depict the passenger line choice behavior under the given transit network defined by the upper-level problem. The passengers consider a deterministic set of attractive lines in their route choice decisions at a stop and board the first bus that arrives at that stop (Spiess and Florian 1989). Recently, by using techniques of ATIS and ITS, analyzing models considering passengers' dynamic travel choices, stochastic traffic conditions and bus arrival time have been extensively studied to better reflect the real-time passengers' travel behavior and bus operating characteristics (Jenelius and Cats 2015; Comi and Nuzzolo 2017).

Similar to the network representation proposed by Nguyen and Pallottino (1988), the transit network is represented by nodes and links, with two adjacent nodes connected by links. It should be mentioned that each link generally has three attributes: travel time, frequency, and capacity (Szeto and Jiang 2014). In this paper, the capacity of vehicles is not taken into consideration, and it is assumed that all passengers can successfully board the bus according to their travel strategy without overcrowding/congestion issues.

Fig. 5 depicts two consecutive transit stops served by k transit lines. The bus stop and transit lines are connected by a pair of boarding and alighting links, which are only adopted for the connectivity purpose. Hence, in order to ensure this purpose, the travel times on these links are set to zero and the frequency of the boarding link is equal to the frequency of the transit line the passenger chooses, whereas the frequency of the alighting link is set to a relatively large number. There is no alighting link for a starting terminal and no boarding link for an ending terminal. Additionally, consecutive stops in a transit line are connected by travel links. The travel time on each link is equal to the sum of in-vehicle travel time and node the stop time at the next stop. The stop time of each terminal is set to zero (Szeto and Jiang 2014). The frequency of each travel link is equal to the frequency of its corresponding transit service.

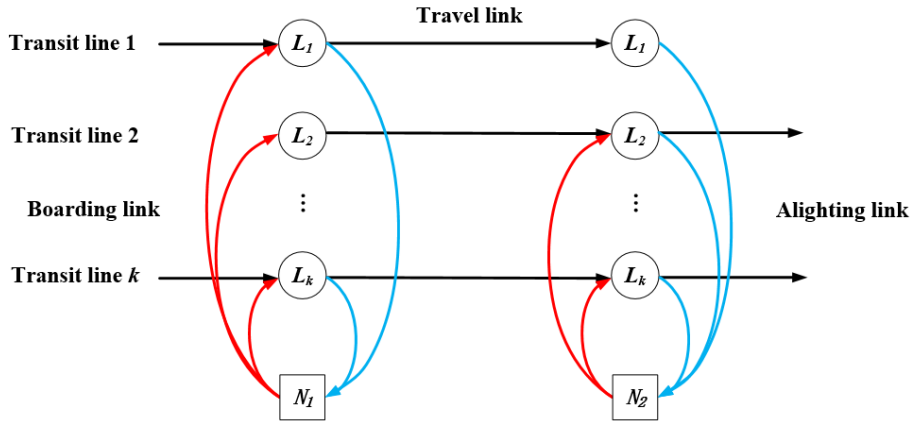


Fig. 5 Illustration of general transit stops.

In this network, a travel strategy is defined as a set of rules that passengers would apply or adhere to. Specifically, in the multimodal transit network $G=(N,A)$ and a

given OD pair w , a strategy is a subset of links in A that represents the possible lines the passenger would choose to minimize their total travel time, which can be termed as a set of attractive lines among all the possible lines between this OD pair. In this process, the following assumptions are made: i) the passenger has complete knowledge of the transit network, including the in-vehicle travel time and the frequency; ii) the passenger would board the first vehicle belonging to his strategy while waiting at a bus stop. Based on the above assumptions, the strategy-based transit assignment model proposed by Spiess and Florian (1989) can be formulated as follows,

$$\min_{\mathbf{v}, \mathbf{w}} z_3 = \sum_{w \in W} \left(\sum_{a \in A} c_a v_a + \sum_{i \in N} \omega_i \right) \quad (18)$$

subject to

$$v_a \leq f_a \omega_i, a \in A, i \in N \quad (19)$$

$$\sum_{a \in A_i^+} v_a = \sum_{a \in A_i^-} v_a + d_i, i \in N \quad (20)$$

$$v_a \geq 0, \omega_i \geq 0, a \in A, i \in N. \quad (21)$$

The objective of the lower-level problem is to minimize the sum of the total in-vehicle travel time and total waiting time between all OD pairs. Constraint (19) is a relaxed constraint that reflects the relationship between link flow v_a , frequency f_a and waiting time ω_i . Specifically, if link a is not a part of the optimal strategy, v_a is zero; otherwise, this constraint verifies with equality. Constraint (20) is the flow conservation equation for each node. Constraint (21) is the non-negativity condition.

5. Solution algorithm for the BLPP

The bi-level model addressed in this paper includes a frequency optimization model and a strategy-based transit assignment problem, which is known as an NP-hard problem (Baaj and Mahmassani 1995). Hence, considering the computational efficiency in practice, the heuristic is more suitable than exact algorithms. In this section, the ABC algorithm (Szeto et al. 2011; Szeto and Jiang 2014; Huang et al. 2016) is

proposed to solve the bi-level problem.

The ABC algorithm is inspired by the intelligent food finding the behavior of honey bees, which can be subsumed to the evolutionary algorithm. It outperforms some other heuristic algorithms (such as the Generic Algorithm) by its inherent local search mechanism. In this algorithm, the bees are classified into three categories: employed bees, onlookers, and scouts. During the food searching process, the employed bee is appointed to explore the available food sources (solutions) until they are exhausted. In details, a local search mechanism is performed in each solution until this solution cannot be improved any more. The onlookers and scouts are in charge of evaluating and searching new solutions respectively.

In this application of the ABC algorithm, each food source is composed of the decision variables in the upper-level problem, which are the bus line set and corresponding frequencies. Note that the set of main bus lines obtained in Set C is considered as the input for the ABC algorithm. Specifically, these lines act as the pool of candidate lines from which the ABC algorithm selects to configure the main bus network. Assume that each line in Set C is identified by a line number. Hence, the main bus network which is composed of lines can be denoted by a fixed length of string. Let N_L denotes the number of main bus lines. Then, the composition of a feasible solution can be represented as shown in Table 2. The detailed algorithmic steps are described as follows.

Table 2 Demonstration of variables.

Variables				
	Main bus lines	Rail	Main Bus	Feeder Bus
Food Source 1	N_L length of string	$f_{R,1}^1, f_{R,1}^2, \dots, f_{R,1}^m$	$f_{B,1}^1, f_{B,1}^2, \dots, f_{B,1}^n$	$f_{F,1}^k, f_{F,1}^k, \dots, f_{F,1}^k$
...
Food Source i	N_L length of string	$f_{R,i}^1, f_{R,i}^2, \dots, f_{R,i}^m$	$f_{B,i}^1, f_{B,i}^2, \dots, f_{B,i}^n$	$f_{F,i}^k, f_{F,i}^k, \dots, f_{F,i}^k$

- Step 1.* Initialization: set the colony size N_c , the number of employed bees N_e , the number of onlooker bees N_o , the number of scout bees N_s , the limit counter L ; set the counter of iterations I to 0, and the maximum number of iterations, I_{\max} .
- Step 2.* Initialization of employed bees: Generate an initial set of food sources. Initialize the limit counter of each food source as 0.
- Step 3.* Employed bee phase: Execute a neighborhood search for each food source that is chosen by the employed bees. In details, input the frequency of each solution into the *lower-level problem* to solve a strategy-based transit assignment subroutine. Input the results of transit assignment to the *upper-level problem* to calculate the objective value and fitness of this food source (as shown in Fig. 6). Evaluate the quality of each food source in terms of the fitness. Replace the current solution if its neighbor solution is better. Otherwise, retain the current solution and increase the limit counter by 1;

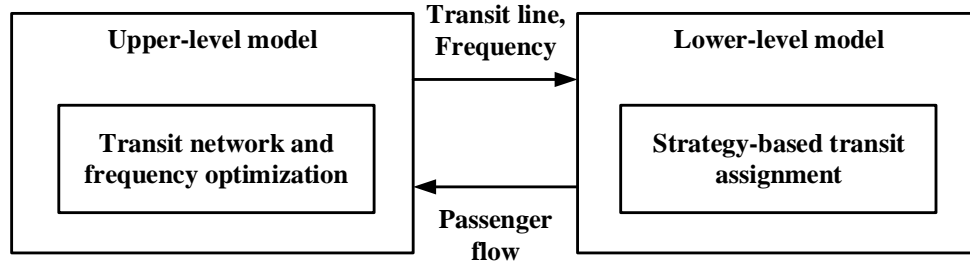


Fig. 6 The framework of the BLPP.

- Step 4.* Onlooker bee phase: Execute the roulette wheel selection in terms of the fitness of each food source to determine which current food source occupied by the employed bee would be chosen by an onlooker bee. Execute a neighborhood search for each food source chosen by onlookers. Evaluate the quality of these food sources. Replace the current solution if its neighbor solution is better. Otherwise, retain the current solution and increase the limit counter by 1;

Step 5. Scout bee phase: For each food source, if it cannot be improved within the maximal trial number limit counter, this food source is treated as a poor one. Then this food source is disregarded and the corresponding employed bee becomes a scout bee, which will search a new food source location randomly and set the limit counter of the new food source as zero.

Step 6. Stop test: Increase the number of iteration, I , by $I = I + 1$. If $I < I_{\max}$, return to Step 3; otherwise, stop.

Remark 1. presented as follows: Assume that the number of variables is D . The food source i of t th iteration is denoted by $X_i^t = [x_{i1}^t, x_{i2}^t, \dots, x_{iD}^t]$, $x_{id} \in (L_d, U_d)$, where L_d and U_d is the lower bound and the upper bound of the search space, respectively. The initial location of the food source i can be obtained by $x_{id} = L_d + \alpha(U_d - L_d)$. In the scout bee phase, the new food source is generated by $v_{id} = x_{id} + \beta(x_{id} - x_{jd})$, where v_{id} is the new food source. α and β is a random variable following the uniform distribution within $[-1, 1]$, which measures the magnitude of the perturbation.”

6. Numerical example

6.1 Small instance

Recall that the main objective of this study is to design bus lines based on a multimodal hub-and-spoke transit network framework, where a newly constructed rail system is taken as the backbone. The proposed model is tested on a rail network with two rail lines and a set of demand nodes which are randomly generated in the service area. The proposed ABC algorithm was coded in MATLAB R2015b and the lower-level problem was solved by CPLEX 12.2 on a personal computer with Intel Core i5-5200 CPU @ 2.20 GHz and 4.00 G RAM.

6.1 Small transit system

Fig. 7(a) depicts the rail network which consists of two rail lines intersecting at node 4. In this example, the radius of the catchment area of a hub rail station is 1 km. The minimum distance between hubs is 2 km. The radius of the catchment area of a

main bus station is 0.3 km and the minimum distance between hubs is 1km (i.e. $d_0^1 = 1 \text{ km}$, $d_0^{1'} = 2 \text{ km}$, $d_0^2 = 0.3 \text{ km}$ and $d_0^{2'} = 1 \text{ km}$). The total passenger demand is 1000 passenger/h. The passenger's value of time, $\lambda = 0.3$. The parameters of the ABC algorithm were set as follows: $N_{Colony\ bee} = 100$, $N_{Employee} = N_{Onlook} = 50$, limit counter $L = 5$, and the maximum number of iterations I_{max} is 1000.

The decision graphs of the proposed two-phase clustering algorithm in Stage 1 are presented in Fig. 7. In Fig. 7(b), 7 nodes are generated as candidate cluster centers. However, only 6 rail stations are selected as hubs in Fig. 7(a). One of the cluster centers is eliminated, which is highlighted by dashed cycles because no rail station can be found in the catchment area within this cluster. Fig. 7(c) shows that 9 cluster centers are identified and all of them are termed as the main bus stations in Fig. 7(a). Fig. 9 shows the convergence trend of the proposed model and solution algorithm. The objection function decreased by about 24% and 20% against the starting solution for each scenario in 1000 iterations.

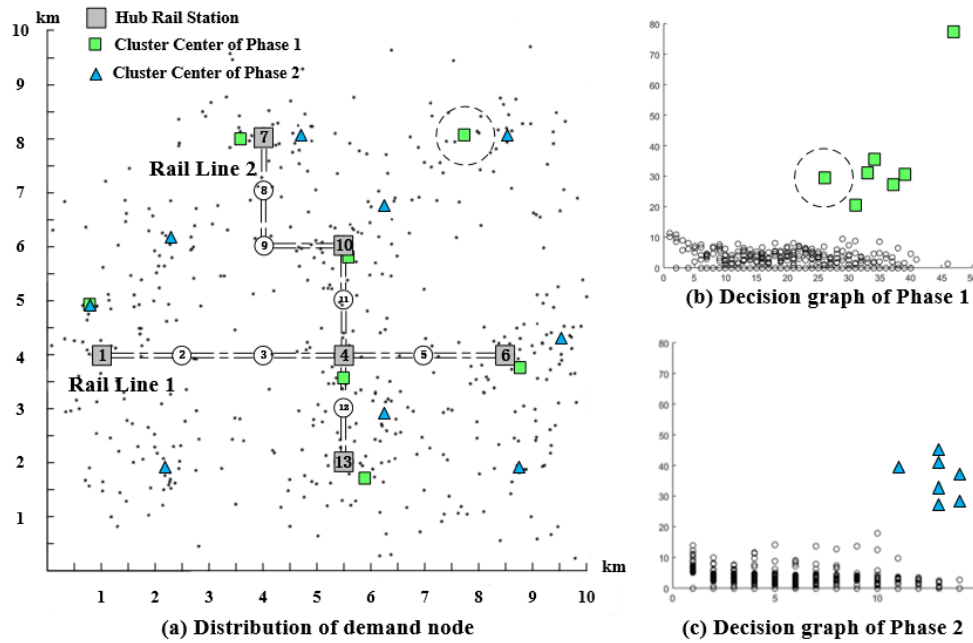


Fig. 7 The example rail network and the distribution of demand nodes.

To illustrate the efficiency of the hub-and-spoke transit network, two scenarios are analyzed. In Scenario 1, all the four criteria for examining the candidate lines are considered simultaneously, including the hub passage constraint, the transit line

overlapping constraint, the total length constraint, and the minimum intersection constraint. In Scenario 2, meanwhile, the hub passage constraint is relaxed, which means that a candidate line can be selected with or without passing hubs. The results of both scenarios are shown in Fig. 8 and Table 3. Without the hub passage constraint, more direct lines are selected in Scenario 2, resulting in a greater average line length (12.36 km) than that of Scenario 1 (11.84 km). In doing so, the number of transfers decreases from 2384 to 1947. However, because of the overlong bus line, the total system in-vehicle travel time in Scenario 2 is greater than that of Scenario 1, resulting in greater operator and user costs in the meantime. Additionally, the computing time for Scenario 1 is about 720 seconds while that for Scenario 2 is about 660 seconds. This is expected because the Scenario 1 has a much smaller solution space than that of Scenario 2, where the hub passage constraint is relaxed, and the number of candidate routes among different OD pairs of Scenario 2 (175 routes) is much more than Scenario 1 (65 routes).

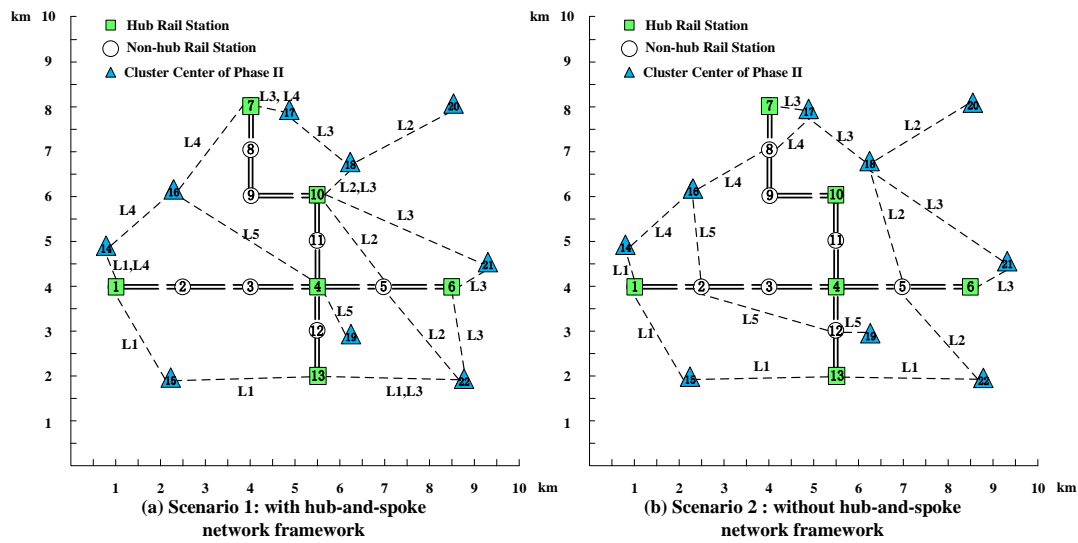


Fig. 8 Transit network with and without hub-and-spoke network framework.

Table 3 Objective function terms for different scenarios.

Objective function value	Operator cost (\$)	User cost (h)	No. of transfers	In-vehicle time (h)	Waiting time (h)	Computing time (s)
-----------------------------	-----------------------	------------------	---------------------	------------------------	---------------------	-----------------------

Scenario 1	240102.1	222914.7	17187.47	2384	16632.2	555.27	920
Scenario 2	259142.1	241848.6	17293.46	1947	16832.25	528.28	660

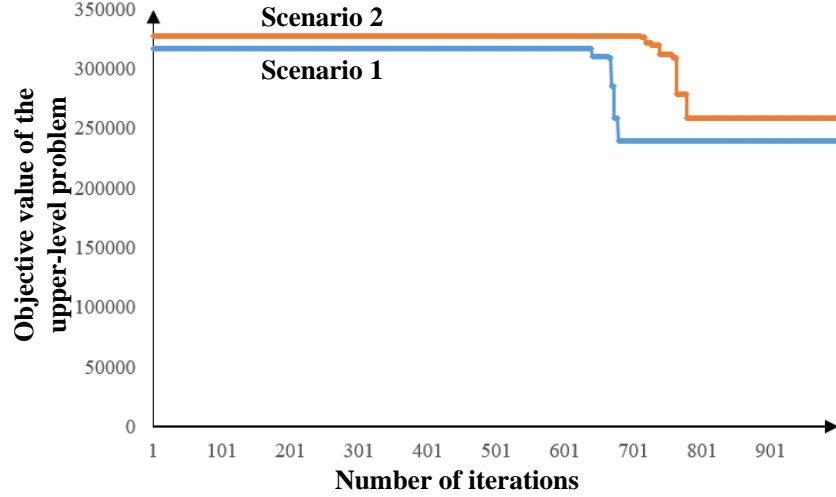


Fig. 9 Convergence trend of the objective value of the upper-level problem.

As it can be seen, the Scenario 1 outperforms the Scenario 2 in the objective value of the upper-level problem. In the perspective of passenger's line choice, the total number of transfer in Scenario 1 is greater than that of Scenario 2, which results in a higher total waiting time. However, the in-vehicle travel cost in Scenario 1 is less than Scenario 2 because of the higher usage of the rail service (see Table 3). This implies a trade-off between traveling directly by bus only and transferring to a rail service in a multimodal transit network, which can be investigated in future studies.

To assess the performance of each transit link, the operating efficiency of link a (denoted as E_a) is introduced as the measurement (Wang and Chen 2012), which is calculated by

$$E_l = \frac{\sum_i \sum_j q_{ij}^l \cdot Length_l}{\sum_{l \in L_R \cup L_B} Length_l} \quad (22)$$

where $Length_l$ is the itinerary length of the transit line l , $l \in L_R \cup L_B$.

The efficiencies of each transit line are presented in Table 4-6. In terms of rail

lines, due to the aggregation of passenger flows, both Rail lines 1 and 2 in Scenario 1 have a better efficiency performance than that of Scenario 2. Another benefit of the consolidation the passenger flow is to reduce the total system travel time. It is calculated that the total system travel time of each scenario is 16632.2 hours and 16832.25 hours respectively, which can be attributed to the higher operation efficiency of the rail system in Scenario 1.

The passenger volumes of each line in Scenario 1 and 2 are presented in Fig. 10. The amount of passenger volume is represented by the darkness of color ranging from 0 to 500. The transfer rates of each rail station are also presented. It is shown that, in Scenario 1, due to the hub passage constraint, passengers intend to transfer at hubs to rail lines in order to minimize their trip time. This statement is also illustrated in the Scenario 2. The transfer rate of Node 1 in Scenario 2 is even higher than that in Scenario 1, which means that the rail lines would become more attractive once it is connected with other transit modes, hence evolution towards a potential MaaS type operation once the integrated information and payment feature are developed to support such a proposition (Blythe et al. 2000). Moreover, the main bus lines with higher itinerary length carry lower passenger volumes resulting in the low operating efficiencies as well. In other words, the construction of hub-and-spoke transit network based on the rail network can efficiently decrease the total travel time of the entire transit system by making efficient use of rapid transit modes.

Table 4 Optimal line volume and frequency of rail lines.

Scenario 1				Scenario 2			
No.	Frequency ^a	Volume ^b	Efficiency ^c	No.	Frequency	Volume	Efficiency
Rail Line 1	12	494	68.01	Rail Line 1	9	425	65.71
Rail Line 2	14	968	133.35	Rail Line 2	8	638	85.14

Note: ^aFrequency is measured by vehicle/h. ^bVolume is measured by passenger/hour. ^cEfficiency is measured by passenger-km/km.

Table 5 Optimal line volume and frequency of main bus lines.

	Line description	Line length ^a	Frequency	Volume	Efficiency
Scenario 1	14-1-15-13	10.4	8	164	20.82
	20-18-10-5-22	14.26	9	246	43.01
	7-17-18-10-21-6-22-13	18.76	9	370	85.09
	17-7-16-14	7.04	8	269	23.20
	16-4-19	8.76	7	135	14.47
Scenario 2	14-1-15-13-22	15.05	7	216	38.56
	20-18-5-22	14.38	7	240	40.93
	7-17-18-21-6	7.17	8	115	9.78
	17-8-16-14	6.74	6	95	7.59
	16-2-12-19	18.46	7	220	48.17

Note: ^aLine length is measured by km.

Table 6 Optimal frequency of feeder bus lines.

Scenario 1				Scenario 2			
No. ^a	Frequency	No.	Frequency	No.	Frequency	No.	Frequency
1	4	8	2	1	3	8	5
2	4	9	3	2	4	9	3
3	2	10	2	3	2	10	3
4	5	11	2	4	5	11	2
5	5	12	2	5	5	12	2
6	5	13	2	6	5	13	2
7	2			7	5		

Note: ^aThe No. of feeder bus refers to the sequence number of rail stations.

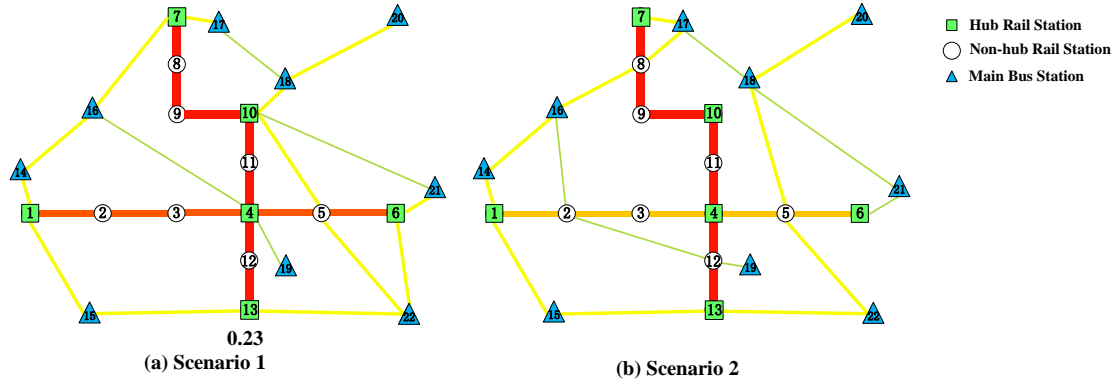


Fig. 10 Passenger volume of each transit line.

6.2 Mandl's benchmark

The proposed model and solution method is then applied in the Swiss network, which is also named as the Mandl's benchmark with 15 nodes and 21 links and widely adopted by previous studies (Baaj et al. 1991; Chakroborty et al. 2002; Bagloee and Ceder 2011; Yan et al. 2013). Different from aforementioned studies, a rail line is constructed across the road network between the nodes with higher passenger demand (the daily OD matrix can be obtained from Mandl (1980)). The link travel times of rail lines are assumed to be half of those of main bus lines. In this experiment, due to the absence of the actual demand distribution, Stage 1 is skipped and the hub location is simply selected based on the amount of the node passenger production/attraction. Taking into account this observation, Node 6 and 10 on this rail line are selected as the candidate hub rail stations, the locations of hubs are analyzed in following three scenarios: (i) Node 10; (ii) Node 6 and (iii) both Node 6 and 10.

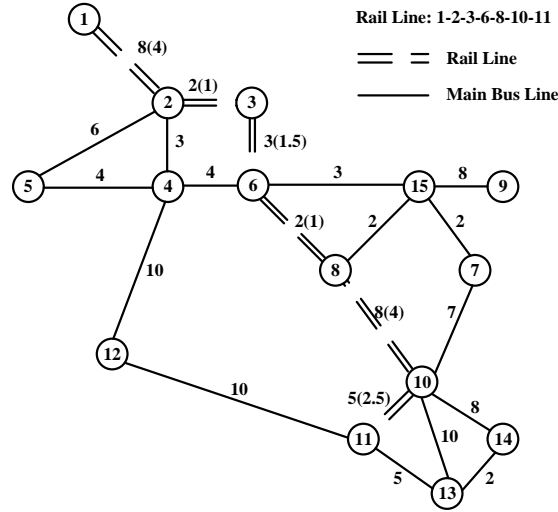


Fig. 11 The modified Swiss road network with a rail line.

The results of the transit network design and optimal line frequency obtained in Stage 2 and 3 are shown in Fig. 12 and Table 7. According to the results, we found that the location of the hub rail station has significant impacts on the main bus line network design and the distribution of passenger demand. In the proposed network framework, the main bus line is restricted to pass at least one hub rail station, which is a strong assumption that may cause overlapping lines especially in a sparse road network. In Scenario 1, the location of the hub rail station (Node 10) spatially lies deviated from the center of the network, causing a significant overlapping between the rail line and the other three main bus lines (Line 1, 2 and 3). This is similar to Scenario 2 where the overlap between rail and main bus lines is reduced and thus the efficiency of the rail line is improved. In the Scenario 3, two rail stations are selected as hubs so that the hub passage constraint is relaxed resulting in a significant increase of passenger flow transfer to the rail line and an improvement of the average efficiency of the transit system. The increasing number of hubs would accordingly extend the solution space and results in a greater computing time (see Table 7). The computing time of Scenario 3 with two candidate hubs is 3.5 hours, which is almost three times than that of Scenario 1 and 2.

Table 7 Optimal line volume and frequency of main bus lines.

	No. of lines	Line Description	Line length	Frequency	Volume	Efficiency	Computing time
Scenario 1	Rail	1-2-3-6-8-10-11	14	13	5399	359.91	1.3 hours
Node 10	1	1-2-3-6-8-10-7	30	6	983	140.40	
	2	5-4-6-8-10-11-13	28	9	693	92.35	
	3	5-4-6-8-10-7	25	9	950	113.07	
	4	9-15-7-10-11-12-4-2-5	51	9	1133	275.13	
	5	9-15-7-10-13-14	29	10	672	92.74	
	6	4-12-11-10-14	33	7	743	116.76	
Scenario 2	Rail	1-2-3-6-8-10-11	14	10	5582	465.16	1.3 hours
Node 6	1	1-2-3-6-15-7	18	9	957	102.50	
	2	5-4-6-8-10-11-13	30	10	840	150.08	
	3	5-4-6-15-7	13	7	674	80.21	
	4	9-15-6-4-2-5	24	6	791	113.07	
	5	9-15-6-4-12-11-13-14	39	7	765	132.07	
	6	4-6-8-10-14	22	7	573	146.93	
Scenario 3	Rail	1-2-3-6-8-10-11	14	13	7053	617.16	3.5 hours
Node 6&10	1	1-2-3-6-15-7	18	6	837	94.11	
	2	5-4-6-8-10-11-13	30	7	714	133.93	
	3	5-4-6-15-7	20	10	615	49.99	

4	9-15-6-4-2-5	24	9	1126	168.95
5	9-15-7-10-13-14	29	6	327	79.74
6	4-12-11-10-14	33	9	748	78.79

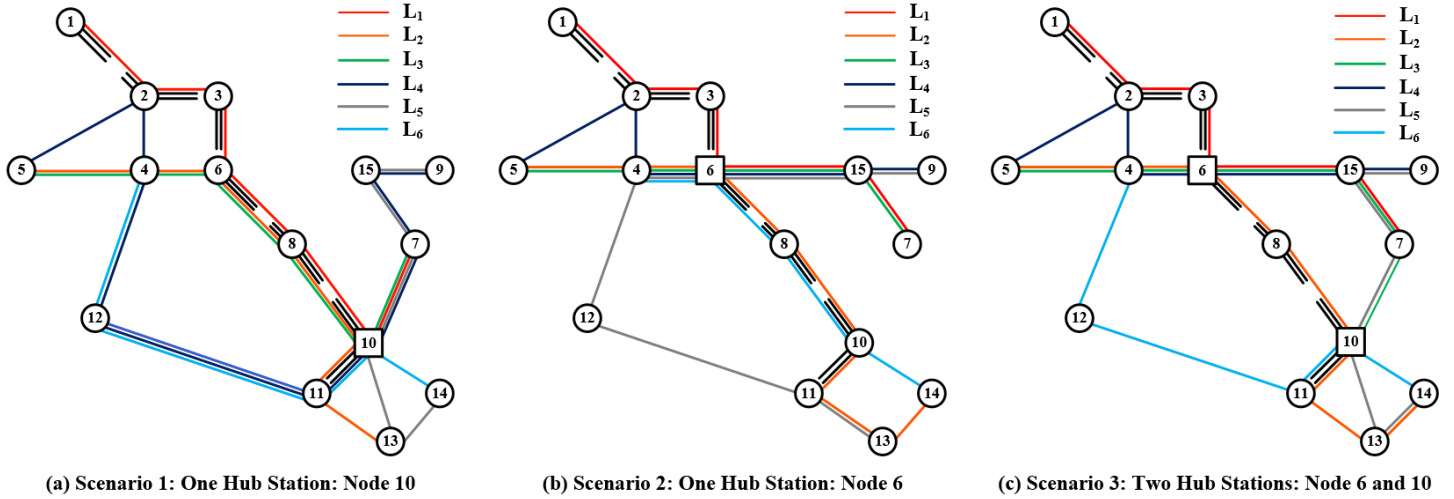


Fig. 12 Main bus network designed in different hub locations.

Due to the similarity of line-generation and line-selection approaches, the proposed methodology is compared to the studies of Mandl (1980), Baaj and Mahmassani (1991), Chakroborty and Dwivedi (2002), Bagloee and Ceder (2011) and Yan et al. (2013). Table 8 shows the results and comparisons between the proposed scenarios and previous studies. It can be seen that the ratios of one transfer and two transfers with respect to the total demand are higher than those of previous studies as in the proposed transit network, the passengers are encouraged to transfer to the rail line to decrease their in-vehicle travel time. As expected, the ratio of no transfer of Scenario 3 with two hubs is the lowest, which evidences the main positive outcome of implementing a hub-and-spoke network framework as aforementioned. Obviously, the construction of the rail line can decrease the total passenger in-vehicle travel time significantly. However, the increasing number of transfers results in an increasing ratio of waiting time and transfer time, which accounts for greater total travel time than reported in previous studies.

The detailed ratio of transfers in different categories of stations is presented in

Table 9. The results show that Scenario 3 outperforms the other two scenarios in the ratio of transferring in hubs resulting from the additional number of hubs. As for the one-hub scenarios, Scenario 2 attracts more passengers to transfer in the hub rail station than Scenario1 indicating that Node 6 is more suitable to be constructed as a hub than Node 10, which is in accordance with the previous observations.

Table 8 Comparison with previous studies.

Solution	Number of lines	% of demand			Total time elements (min)			
		Zero transfer	One transfer	Two transfers	Travel time	In-vehicle time	Out-of-vehicle time	Transfer time
Mandl (1980)	4	69.94	29.93	0.13	219,094	177,400	18,194	23,500
Baaj & Mahmassani (1991)	6	80.99	19.01	0	217,954	180,356	22,804	14,800
Chakroborty & Dwivedi (2002)	7	89.15	10.85	0	-	-	-	-
Bagloee & Ceder (2011)	12	83.66	15.21	0.95	202,255	167,198	24,591	10,465
Yan et al. (2013)	6	89.93	10.07	0	-	-	-	-
Current study:								
Scenario 1	6	74.07	24.40	1.53	164,290	124,552	29,047	10,691
Scenario 2	6	78.08	20.84	1.08	146,792	113,364	26,265	7,163
Scenario 3	6	62.88	36.41	0.71	155,427	113,754	29,891	11,782

Table 9 Ratio of transfers in different categories of stations.

	% of total transfers		
	Transfer in hub station	Transfer in non-hub station	Transfer in bus station
Scenario 1	41.37	52.11	6.52
Scenario 2	57.16	33.13	9.71
Scenario 3	74.84	12.64	12.52

6.3 Nanjing rail system

To illustrate the effectiveness of the proposed approach in a large-scale real network, this study has collected a big amount of field data of public transport system in Nanjing, China. The data include: the total urban area (243 square kilometers), and the smart card data in public transit system (including both rail and bus). There are 3 rail transit lines and 113 stops in the urban area of Nanjing (see Fig. 13). The total rail line length is 121.75 km and the daily average passenger demand is 1.5 million. As illustrated in Section 6.3, the increase of hub rail stations results in a broader solution space and a more extensive candidate route set. A large-scale real transit network causes overwhelming computational burden. Thereby, the partition of the network can arbitrarily reduce its computational burden, which makes the proposed method suitable for practical implementations. In this study, the study area is first partitioned into nine zones, while the clustering of hub location is performed separately in each zone. Note that the distribution of the passenger demand is represented by boarding data collected by the automated fare collection system at each rail and bus station.

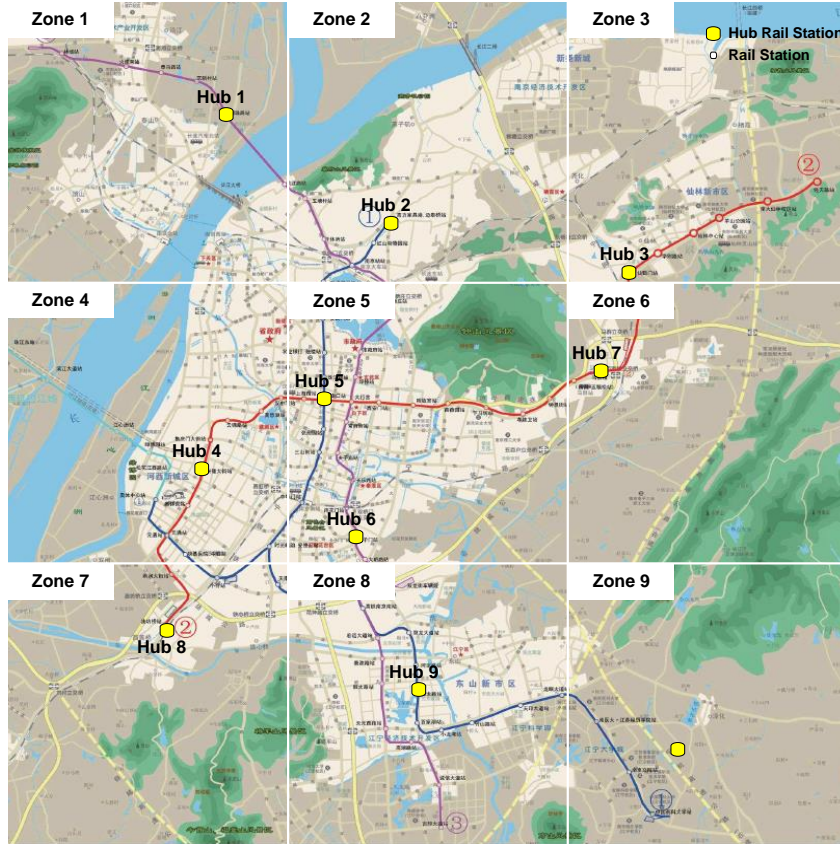


Fig. 13 Rail system in Nanjing.

As shown in Fig. 14, the decision paragraphs of each zone are presented while at least one hub is selected in each zone. Note that the cluster center of zone 9 is not located at the rail station. This case is simple because no rail station is suitable to be a hub in the catchment area within this cluster as a result of the unevenly distributed passenger demand. To better illustrate the decision process of the proposed method, a detailed decision paragraphs and resulting distribution of hub rail and main bus distribution of Zone 5 are presented in Fig. 15, located in the downtown area of Nanjing. In Zone 5, two rail stations are selected as hubs according to the decision graph Fig. 15(b), both of which are located in the area with higher demand density. In Phase 2 of the clustering stage Fig. 15(c), 70 main bus stations are generated which are indicated as black dots in Fig. 15(a).

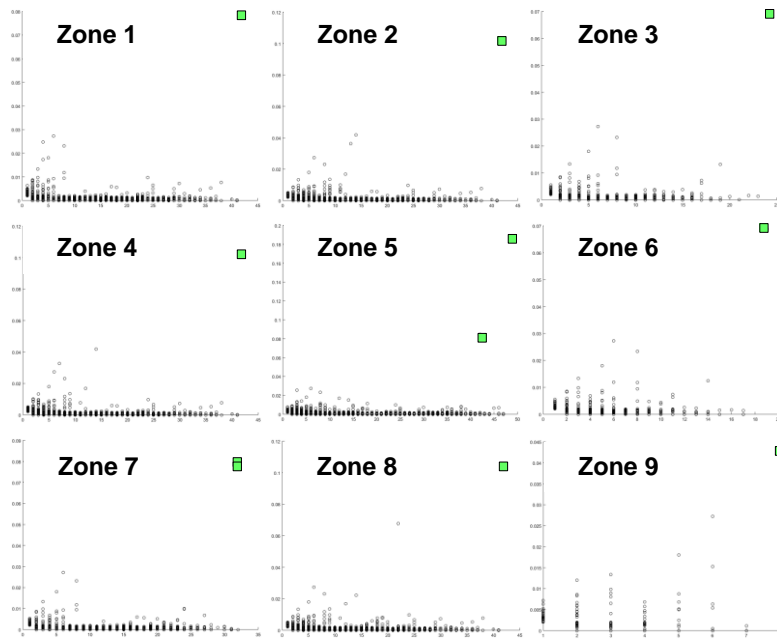


Fig. 14 Decision paragraphs of each zone.

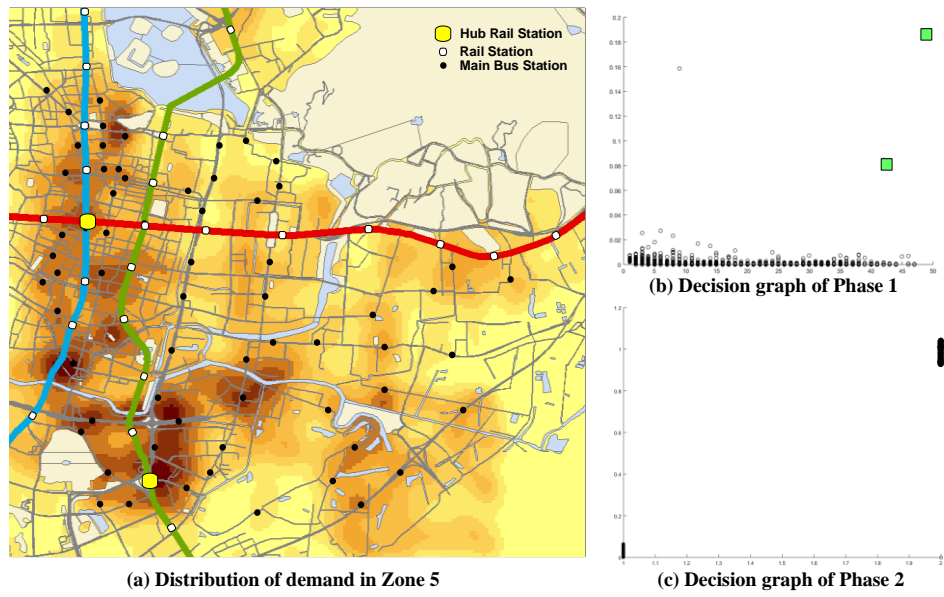


Fig. 15 Distribution of hub rail and main bus stations in Zone 5.

Table 10 Distributions of OD volume, distances and coefficient of direction between hubs.

Hub	Item	1	2	3	4	5	6	7	8	9
1	Volume ^a	0	49	16	18	579	33	20	4	27
	Distance ^b	0.00	9.98	29	24.66	11.01	19.73	20.01	31.65	26.17
	Coefficient ^c	-	1.38	1.8	1.74	1.22	1.18	1.16	1.62	1.16
2	Volume	17	0	3	59	2124	10	14	25	11
	Distance	9.98	0.00	24.02	14.94	8.28	14.27	19.33	21.93	20.71
	Coefficient	1.38	-	2.61	1.24	1.14	1.18	1.95	1.26	1.17
3	Volume	2	0	0	10	264	0	53	6	4
	Distance	29	24.02	0.00	22.41	15.74	20.15	4.69	29.41	26.59
	Coefficient	1.8	2.61	-	1.25	1.49	1.81	1.17	1.34	1.46
4	Volume	1	10	13	0	294	2	12	24	0
	Distance	24.66	14.94	22.41	0.00	6.66	13.11	17.72	6.99	19.55
	Coefficient	1.74	1.24	1.25	-	1.23	2.18	1.15	1.24	1.89
5	Volume	3	48	39	21	0	3	23	6	3
	Distance	11.01	8.28	15.74	6.66	0.00	6.44	11.05	13.66	12.89
	Coefficient	0.96	1.14	1.49	1.23	-	1.21	1.06	1.33	1.17
6	Volume	3	2	4	1	34	0	2	0	8
	Distance	19.73	14.27	20.15	13.11	6.44	0.00	15.46	20.11	6.44
	Coefficient	1.18	1.18	1.81	2.18	1.21	-	1.39	2.57	1.11
7	Volume	2	12	252	33	711	9	0	24	6
	Distance	20.01	19.33	4.69	17.72	11.05	15.46	0.00	24.72	21.91
	Coefficient	1.16	1.95	1.17	1.15	1.06	1.39	-	1.31	1.53
8	Volume	6	41	29	295	893	4	19	0	10

	Distance	31.65	21.93	29.41	6.99	13.66	20.11	24.72	0.00	31.16
	Coefficient	1.62	1.26	1.34	1.24	1.33	2.57	1.31	-	3.66
9	Volume	3	5	5	4	93	16	2	0	0
	Distance	26.17	20.71	26.59	19.55	12.89	6.44	21.91	31.16	0.00
	Coefficient	1.16	1.17	1.46	1.89	1.17	1.11	1.53	3.66	-

Note: ^aVolume is measured by passenger/hour. ^bDistance is measured by km. ^cCoefficient of direction.

7. Conclusions

In this paper, a multimodal transit network design method is proposed and investigated. Under the background of the newly introduced rail system, this method focuses on the reconfiguration of bus services (including main and feeder bus services) in a hub-and-spoke framework, which would increase the network service efficiency. Understanding such efficiencies and interchanges will help city authorities better design elements of a MaaS transport provision.

The proposed MTNDP is addressed at both the strategic and tactic levels. In the first stage, a cluster-based HLP is solved by a fast and efficient cluster algorithm which is adopted to decide the number of clusters intuitively. A multimodal transit network including main and feeder bus lines is designed hierarchically. For the main bus network, the trade-offs between the user's and operator's view of the bus line are considered separately. Four practical feasibility constraints are used to filter these lines to obtain the final bus network. Comparative studies with respect to hub passage constraint have further demonstrated that the application of hub-and-spoke framework can efficiently decrease the total travel time on the transit system as well as the operator's cost. To obtain the optimal frequency of each transit line, a bi-level model is formulated in the third stage. The upper-level aims to minimize the total system cost including the operation cost and the user's travel cost. The lower-level problem is a transit assignment problem based on the optimal strategy concept.

The network design method developed in this study can be used to optimize the multiple bus line deployment if the city plans to construct/upgrade the rail transit system.

Several potential enhancements could be considered in future works. First, the feasibility of the proposed methodology can be further testified in an at scale road and rail network. The distribution of the passenger demand can also be better reflected by the application of data fusion techniques. Second, the proposed BLPP is considered a leader-follower game from the perspective of the operator and the passenger. Actually, the transportation authority also plays an important role in the MTNDP, which results in a three-party game. Third, advanced spatial partition techniques can be adopted considering characteristics of a realistic transit system.

Acknowledgement

This study is supported by the Natural Science Foundation of Jiangsu Province (No. BK20150603 and BK20160676), the Postgraduate Research and Innovation Plan Project in Jiangsu Province (KYLX16_0277). Further support to foster academic engagement between SEU and Newcastle University was provided by the EPSRC/NSFC project LC Transforms: Low Carbon Transitions of Fleet Operations in Metropolitan Sites Project (EP/N010612/1).

References

- Arbex, R. O., and C. B. da Cunha. 2015. "Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm." *Transportation Research Part B* 81:355-76.
- Aykin, T. 1995. "Networking policies for hub-and-spoke systems with application to the air transportation system." *Transportation Science* 29 (3):201-21.
- Baaj, M. H., and H. S. Mahmassani. 1991. "An AI-based approach for transit route system planning and design." *Journal of Advanced Transportation* 25 (2):187-209.

- Baaj, M. H., and H. S. Mahmassani. 1995. "Hybrid route generation heuristic algorithm for the design of transit networks." *Transportation Research Part C* 3 (1):31-50.
- Badia, H., M. Estrada, and F. Robusté. 2014. "Competitive transit network design in cities with radial street patterns." *Transportation Research Part B* 59:161-81.
- Bagloee, S. A., and A. A. Ceder. 2011. "Transit-network design methodology for actual-size road networks." *Transportation Research Part B* 45 (10):1787-804.
- Blythe, P., T. Rackliff, R. Holland, and J. Mageean. 2000. "ITS applications in public transport: improving the service to the transport system user." *Journal of Advanced Transportation* 34 (34):325-45.
- Blythe, P.T. 2016. "Mobility as a Service – the potential rewards and challenges from a UK perspective." In *24th Intelligent Transport Systems World Congress, Special Session SIS04: Regulation as an enabler for positive transportation change*. Melbourne, Australia.
- Bordagaray, M., L. Dell'Olio, A. Ibeas, and P. Cecín. 2014. "Modelling user perception of bus transit quality considering user and service heterogeneity." *Transportmetrica A: Transport Science* 10 (8):705-21.
- Campbell, J. F. 1996. "Hub location and the p-hub median problem." *Operations Research* 44 (6):923-35.
- Cancela, H., A. Mauttone, and M. E. Urquhart. 2015. "Mathematical programming formulations for transit network design." *Transportation Research Part B* 77:17-37.
- Ceder, A. A. 2007. *Public Transit Planning and Operation: Theory, Modeling and Practice*. Butterworth-Heinemann, Oxford, UK.: Elsevier.
- Chakroborty, P., and T. Wivedi. 2002. "Optimal route network design for transit systems using genetic algorithms." *Engineering Optimization* 34 (1):83-100.

- Chang, H., D. Park, S. Lee, H. Lee, and S. Baek. 2010. "Dynamic multi-interval bus travel time prediction using bus transit data." *Transportmetrica* 6 (1):19-38.
- Chandra, S., and L. Quadrioglio. 2013. "A model for estimating the optimal cycle length of demand responsive feeder transit services." *Transportation Research Part B* 51:1-16.
- Cipriani, E., S. Gori, and M. Petrelli. 2012. "Transit network design: A procedure and an application to a large urban area." *Transportation Research Part C* 20 (1):3-14.
- Comi, A., and A. Nuzzolo. 2017. "A Dynamic Strategy-based Path Choice Modelling for Real-time Transit Simulation." *In Modelling Intelligent Multi-Modal Transit Systems*, 152-173. CRC Press.
- Daganzo, C. F. 2010. "Structure of competitive transit networks." *Transportation Research Part B* 44 (4):434-46.
- de Sá, E. M., I. Contreras, and J. Cordeau. 2015. "Exact and heuristic algorithms for the design of hub networks with multiple lines." *European Journal of Operational Research* 246 (1):186-98.
- Farahani, R. Z., M. Hekmatfar, A. B. Arabani, and E. Nikbakhsh. 2013. "Hub location problems: A review of models, classification, solution techniques, and applications." *Computers & Industrial Engineering* 64 (4):1096-109.
- Farahani, R. Z., E. Miandoabchi, W. Y. Szeto, and H. Rashidi. 2013. "A review of urban transportation network design problems." *European Journal of Operational Research* 229 (2):281-302.
- Gelareh, S. 2008. *Hub Location Models in Public Transport Planning*: VDM Publishing.
- Gelareh, S., and S. Nickel. 2011. "Hub location problems in transportation networks."

- Transportation Research Part E* 47 (6):1092-111.
- Huang, D., Z. Liu, L. Pan, and J. Chen. 2016. "Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure." *Transportation Research Part E* 96:1-19.
- Iyigun, C. 2013. "The planar hub location problem: a probabilistic clustering approach." *Annals of Operations Research* 211 (1):193-207.
- Jenelius, E., and O. Cats. 2015. "The value of new public transport links for network robustness and redundancy." *Transportmetrica A: Transport Science* 11 (9):819-35.
- Klincewicz, G. 1991. "Heuristics for the p-hub location problem." *European Journal of Operational Research* 53 (1):25-37.
- Kriegel, H., P. Kröger, J. Sander, and A. Zimek. 2011. "Density-based clustering." *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery* 1 (3):231-40.
- Li, Z. C., W. H. K. Lam, and S. C. Wong. 2009. "Optimization of a bus and rail transit system with feeder bus services under different market regimes." In *Transportation and Traffic Theory 2009:Golden Jubilee*, 495-516. Springer US.
- Li, Z. C., W. H. K. Lam, and S. C. Wong. 2011. "On the allocation of new lines in a competitive transit network with uncertain demand and scale economies." *Journal of Advanced Transportation* 45 (4):233-51.
- Mandl, C. E. 1980. "Evaluation and optimization of urban public transportation networks." *European Journal of Operational Research* 5 (6):396-404.
- Nguyen, S., and S. Pallottino. 1988. "Equilibrium traffic assignment for large scale transit networks." *European Journal of Operational Research* 37 (2):176-86.
- Nickel, S., A. Schöbel, and T. Sonneborn. 2001. "Hub location problems in urban traffic

- networks." *In Mathematical Methods on Optimization in Transportation Systems*, 95-107. Springer.
- Nuzzolo, A., and A. Comi. 2016. "Advanced public transport and intelligent transport systems: new modelling challenges." *Transportmetrica A: Transport Science* 12 (8), 674-699.
- O'Kelly, M. E. 1986. "The location of interacting hub facilities." *Transportation Science* 20 (2):92-106.
- O'Kelly, M. E. 1992. "A clustering approach to the planar hub location problem." *Annals of Operations Research* 40 (1):339-53.
- Rodriguez, A., and A. Laio. 2014. "Clustering by fast search and find of density peaks." *Science* 344 (6191):1492-6.
- Saidi, S., S. C. Wirasinghe, L. Kattan, and S. Esmailnejad. 2017. "A generalized framework for complex urban rail transit network analysis." *Transportmetrica A: Transport Science* 13 (10):874-92.
- Sang, Y. L., K. E. Yoo, and Y. Park. 2014. "A continuous connectivity model for evaluation of hub-and-spoke operations." *Transportmetrica A: Transport Science* 10 (10):894-916.
- Schöbel, A. 2012. "Line planning in public transportation: models and methods." *OR Spectrum* 34 (3):491-510.
- Spiess, H., and M. Florian. 1989. "Optimal strategies: A new assignment model for transit networks." *Transportation Research Part B* 23 (2):83-102.
- Sung, C. S., and H. W. Jin. 2000. "A tabu-search-based heuristic for clustering." *Pattern Recognition* 33 (5):849-58.
- Sung, C. S., and H. W. Jin. 2001. "Dual-based approach for a hub network design problem under non-restrictive policy." *European Journal of Operational*

Research 132 (1):88-105.

Szeto, W. Y., Y. Wu, and S. C. Ho. 2011. "An artificial bee colony algorithm for the capacitated vehicle routing problem." *European Journal of Operational Research* 215 (1):126-35.

Szeto, W. Y., and Y. Jiang. 2014. "Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach." *Transportation Research Part B* 67:235-63.

Szeto, W. Y., Y. Jiang, D. Z. W. Wang, and A. Sumalee. 2015. "A sustainable road network design problem with land use transportation interaction over time." *Networks and Spatial Economics* 15 (3):791-822.

Szeto, W. Y., and A. B. Wang. 2016. "Reliable network design under supply uncertainty with probabilistic guarantees." *Transportmetrica A: Transport Science* 12 (6):1-33.

Wang, Z., and Y. Chen. 2012. "Development of location method for urban public transit networks based on hub-and-spoke network structure." *Transportation Research Record: Journal of the Transportation Research Board* (2276):17-25.

Yan, Y., Z. Liu, Q. Meng, and Y. Jiang. 2013. "Robust optimization model of bus transit network design with stochastic travel time." *Journal of Transportation Engineering* 139 (6):625-34.

Yang, T. 2010. "A two-stage stochastic model for airline network design with uncertain demand." *Transportmetrica* 6 (3):187-213.

Yu, B., H. Zhu, W. Cai, N. Ma, Q. Kuang, and B. Yao. 2013. "Two-phase optimization approach to transit hub location—the case of Dalian." *Journal of Transport Geography* 33:62-71.

Appendix 1

The tentative frequency is an auxiliary variable which is applied to manage the insertion sequence of links in the route. The vehicle load factor is a coefficient that represents crowding, which must be less than a predefined maximum value in order to guarantee the service quality for passengers by limiting the crowding effect inside the vehicle (Arbex et al., 2015). Thus, the frequency can be calculated using the point-check period max load method proposed by Ceder (1987):

$$f_l = \frac{v_{\max}^l}{LF_{\max}^l \cdot k_{cap}^l} \leq f_{\max}^l$$

where f_l is the frequency of line l . f_{\max}^l is the maximum frequency of line l . v_{\max}^l is largest link volume of line l . LF_{\max}^l is the maximum load factor of line l . k_{cap}^l is the bus capacity.

In this work, the line frequency is optimized by the proposed BLPP. And the tentative frequency calculated by Eq. 4 is only adopted to manage the link insertion procedures. Additionally, we assume that the LF_{\max}^l , LF_{\min}^l , k_{cap}^l are predetermined parameters. The construction of lines is started from the destination terminal and the specific procedures are presented as follows:

Step 1. For each terminal, the incoming link is selected if the link volume is greater than a predetermined minimum value, v_{\min} ; Add the incoming line in line l .

Step 2. Use Eq. 4 to calculate the tentative frequency of line l ;

Step 3. Expand line l by inserting the link with the highest volume starting from the node reached by line l ;

Step 4. Calculate the load factor of the incoming link, $LF = \frac{v_{\max}^l}{f_l \cdot k_{cap}^l}$;

Step 5. If the LF of the incoming link is lower than LF_{\min}^l , stop the insertion of line l ; otherwise, go to Step 3;